A modified bootstrap for importance sampled data[1]

R.L. Harrison, J.-S. Kim, T.K. Lewellen

Dept. of Radiology, University of Washington, Seattle, WA 98195-6004

Abstract

We are testing a modified bootstrap technique for importance sampled data from SimSET (a simulation system for emission tomography). The bootstrap allows us to simultaneously produce multiple raw data sets from a single simulation while at the same time reducing the weight variation caused by importance sampling. This combination may greatly reduce the CPU time required to produce the multiple image realizations needed for ROC studies. Initial testing indicates that the mean and variance of medium and high count bins in the raw data and high-activity regions of interest (ROIs) are reproduced relatively accurately.

I. Introduction

We are developing a modified bootstrap technique to apply to emission tomography simulations using SimSET [1].

Despite the use of importance sampling (IS) [2] and the ever-increasing speed of computers, many problems remain unwieldy or intractable to study using simulation. One such problem is the generation of multiple realizations of the same image for ROC studies. These studies often require hundreds of image realizations from each of several different activity distributions. Such data sets can take months or even years of CPU time to generate.

The problem of generating multiple samples from a probability distribution when only one data set is available, or when multiple data sets are difficult to produce, is often attacked using the bootstrap technique [3, 4]. In this technique, the available data set is used as an estimate of the underlying distribution. New data sets are generated by sampling data points from the original data with replacement.

We have attempted to modify the bootstrap to ameliorate a problem inherent to the use of importance sampled data: because the events have different weights, the variance has different properties than that of the equivalent analog (non-importance sampled) distribution. The idea underlying our modified bootstrap is that by sampling events with frequency proportional to their weight, we can reduce or eliminate the variation in weight. One can imagine several ways of realizing this goal: the one we chose is described below in Section II.d.

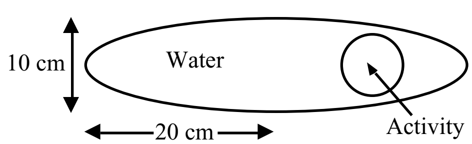

Figure 1: The simulated phantom.

We have tested this technique by generating 99 positron emission tomography (PET) data sets and comparing them to 99 data sets from an analog simulation, both in data space and image space. The results raise as many questions as they answer, but give substantial cause for optimism.

II. Methods

A. Simulation Setup

We performed 99 analog simulations of 4 million decays each. We simulated an elliptical cylinder of water, 40 cm major axis, 10 cm minor axis, and 3 cm axially. A 6 cm diameter circular cylinder of activity was centered on the major axis, 10 cm from the center of the phantom (Figure 1). No collimators or detectors were simulated; instead, all photons reaching a target cylinder with energy greater than 400 keV were considered detected. The target cylinder was centered at the same point as the elliptical cylinder, with 90 cm diameter and 3 cm axial extent.

Coincident events were binned into a 64 by 64 distance-angle array, with the distance bins spanning -20 cm to 20 cm.

B. Bootstrap Implementation/Importance Sampled Simulation

We created 99 data sets using a modified bootstrap technique and the data from one simulation. We used the same simulation setup as described above, except with SimSET's stratification, forced detection, and forced non-absorption features turned on.

Often when bootstrap sampling is applied, it is used to create multiple realizations with N detected event from a single simulation or scan with N detected events [5]. We thought that the resulting overlap in the detected events from data set to data set might cause significant correlation between artifacts in the images reconstructed from the resulting data sets. For this reason, we chose to simulate 664,230 decays—using a short training run, we determined this would produce an equivalent number of detected events to ten 4-million decay analog runs.

Our bootstrap technique sampled from this simulation on-the-fly. Each time a detected event was produced, we sampled a random number, R, for each of the 99 data sets from the Poisson distribution with parameter

$$p = \min\left(0.1\frac{w_{event}}{w_{mean}}, 0.5\right) \qquad (1)$$

where w_{event} is the weight of the current event, w_{mean} is the average event weight, and the factor 0.1 is to compensate for the fact that we were simulating ten times as many decays as needed for one data set. The maximum value of 0.5 is set to keep the Poisson parameter small, so that a event does not appear in too many of the data sets. When the sampled R is 0,

[1] Supported by PHS grant CA42593

the event is not binned in that data set. Otherwise the event is binned with its weight adjusted to

$$w_{output} = \frac{R * w_{event}}{p} \qquad (2)$$

In a more typical bootstrap, w_{output} would not include the factor R; instead, R events would be generated. We chose to create a single event to simplify the modifications needed to the SimSET software. In the long run, however, we believe it may be better to generate R separate events.

C. Image Reconstruction

We reconstructed images from the analog and bootstrap simulation data sets using filtered backprojection with a Hamming window, cutoff frequency 0.6.

D. Statistical Analysis in Data Space

The 99 data sets from the analog simulations were compared to the 99 data sets from the bootstrap simulation using the Student t-test and the f-test [6]. We applied these tests bin-by-bin, and grouped the results according to the mean number of counts in the analog simulation data sets.

E. Region-of-Interest Analysis in Image Space

Two square ROIs, 5 by 5 pixels (3.1 cm by 3.1 cm) were placed on each image. The first ROI was centered in the cylinder of activity, the other in a zero-activity area on the other side of the phantom. We computed the sample mean and variance for the analog images and the bootstrapped images. We then computed Student t-test and f-test statistics.

III. RESULTS

A. Simulation Run Statistics

Some simulation efficiency statistics are given in Table 1. The bootstrap/IS method significantly reduced the cost per realization—though we must note that the situation we simulated is one where the IS can help a great deal, something that is not always true.

Table 1
Simulation Efficiency Statistics

	average analog simulation	average bootstrap realization
CPU time (seconds)	1320	7.8
Total detected counts	19,905	22,385
Total detected weight	938,647	916,355
Total detected weight-squared	44.26 million	39.62 million
Quality factor (QF)	1.0	0.95
$\frac{\text{counts} * \text{QF}}{\text{CPU\_time}}$	15.1	2752

B. Data Space Analysis

The results of the Student t-test and the f-test applied bin-by-bin are shown in Tables 2 and 3. The last four columns show the percentage of analog-bootstrap bin pairs with $p > 0.5$, 0.05, 0.01 and ≤ 0.01—the last three being of interest as common levels for the rejection of the null hypothesis. No results are shown for 1223 bins for which one data sets (analog and/or bootstrap) had no counts in any realization. (The analog data sets had 659 such bins, bootstrap 1191.) The rest of the results are sorted by the mean number of counts in the analog data bin. The line for average counts between 0 and 1 should, perhaps, have been omitted, as many of these bins had a total, over all the realization, of 10 or less events. We usually wouldn't apply these tests to Poisson data with less than 20-30 counts.

However, even in the high count bins, we are seeing significantly different means and variances for a high number of bins. The distributions are close, as shown in the fact that there are also many p-values greater than 0.5 and 0.05. However, overall the distribution of data is clearly different for the analog and bootstrap realizations.

Table 2
Bin-by-Bin T-test Comparison

mean counts	# of bins	% p > 0.5	% p > 0.05	% p > 0.01	% p ≤ 0.01
0	1223	-	-	-	-
0 to 1	2209	16.3	48.2	60.8	39.3
1 to 5	97	35.1	69.1	81.4	18.6
5 to 10	67	37.3	71.6	80.6	19.4
10 to 30	185	22.2	58.9	77.3	22.7
30 to 50	183	20.8	54.6	70.0	30.1
50 to 70	132	15.2	49.2	59.9	40.2

Table 3
Bin-by-Bin F-test Comparison

mean counts	# of bins	% p > 0.5	% p > 0.05	% p > 0.01	% p ≤ 0.01
0	1223	-	-	-	-
0 to 1	2209	8.1	25.3	34.2	65.8
1 to 5	97	23.7	70.1	82.5	17.5
5 to 10	67	26.9	85.1	95.5	4.5
10 to 30	185	37.8	79.5	90.3	9.7
30 to 50	183	26.2	78.7	90.2	9.9
50 to 70	132	30.3	81.1	95.5	4.6

C. Image Space Analysis

Table 4 shows the sample mean and variance for the two image ROIs. For both ROIs, the means are very close. The sample variance make clear, however, that this was a matter of luck. The variances for the ROI with activity have an f-test p-value of 0.29, so they are also reasonably close. However, the background variances are significantly different (p = 0.0004).

Table 4
ROI Mean and Variance

	Analog mean	Analog variance	Bootstrap mean	Bootstrap variance
Active ROI	707.2	150.2	707.8	121.3
Background ROI	2.18	73.39	0.19	35.6

IV. Discussion

This is preliminary work, and as such raises far more questions than it answers. The modified bootstrap technique helps to speed up the generation of multiple realizations tremendously, but we have not shown that these realizations are close enough to truly independent realizations to allow the technique to be used.

However, there are a number of ways we can address the problems. A better way to assign the number of decays for the bootstrap simulation would help—our use of a short training run for that purpose will have propagated the noise from the short run through our entire data set. There are other ways to implement the bootstrap algorithm that might help. And, given the tremendous speed-up we achieved, we could experiment with generating more events.

We do not really expect the bootstrap to perfectly reproduce the bin-by-bin statistics shown above. As seen above the performance on image statistics can be very good even in situations where the underlying data sets do not perfectly represent the true distributions. We plan more tests for this technique in image space. The ultimate test will be the performance of the bootstrap on ROC data sets.

Finally, we note that SimSET is available free of charge for non-commercial use. However, the bootstrap algorithm is not in the distributed package, and will not be until we have carried this investigation considerably further. Contact simset@u.washington.edu for more information.

VI. References

[1] T. K. Lewellen, R. L. Harrison, and S. Vannoy, “The SimSET program,” in *Monte Carlo Calculations in Nuclear Medicine*, *Medical Science Series*, M. Llungberg, S.-E. Strand, and M. A. King, Eds. Bristol: Institute of Physics Publishing, 1998, pp. 77-92.

[2] D. R. Haynor, R. L. Harrison, and T. K. Lewellen, “The use of importance sampling techniques to improve the efficiency of photon tracking in emission tomography simulations,” *Med Phys*, vol. 18, pp. 990-1001, 1991

[3] Efron, B., *The Jackknife, the Bootstrap, and other Resampling Plans*. 1982, Philadelphia: Society of Industrial and Applied Mathematics.

[4] Politis, D., Romano, J., and Wolf, M., *Subsampling.* 1999, New York: Springer.

[5] Haynor, D. and Woods, S., *Resampling estimates of precision in emission tomography*. IEEE Trans Med Imag, 1989. **8**(4): p. 337-343.

[6] Press, *Numerical Recipes in C, 2nd ed*. Numerical Recipes in C, 2nd ed. 1992, Cambridge: Cambridge University Press.

.